CS5691 : Pattern Recognition and Machine Learning

Assignment 3 Report

Group No. 23

Group Members :
Aakriti Budhraja
(CS18S009)
Madhura Pande
(CS17S031)
Sadbhavana Babar
(CS18S029)

Course Instructor:
Prof. C. Chandra Sekhar
Professor, Dept. of CSE,
Indian Institute of
Technology, Madras

September 19, 2020

1 Dataset1: Sequential Pattern Classification

We applied Hidden Markov Model(HMM) on this dataset.

1.1 Hidden Markov Model

Hidden Markov Model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (i.e. hidden) states. There are 3 basic types of problems which HMM tries to solve are:

- The Evaluation Problem: Given an HMM model λ and a sequence of observations $O = O_1, O_2, ...O_T$, what is the probability that the observations are generated by the model, ie., $P(O|\lambda)$.
- The Decoding Problem: Given an HMM model λ and a sequence of observations $O = O_1, O_2, ...O_T$, what is the most likely state sequence in the model that produced the observations?
- The Learning Problem: Given an HMM model λ and a sequence of observations $O = O_1, O_2, ...O_T$, how to determine model parameters A, b, Λ which maximizes $P(O|\lambda)$.

We mostly used the first and the third points to solve our given problem of pattern classification. The third problem is solved by Baum-Welch method of parameter estimation and first problem is solved by forward variable method.

1.2 Our Approach

We have 2 types of datasets which are as follows:

- On-line handwritten character data
- Spoken digit data (Isolated digits)

1.3 Observations

1.3.1 On-Line Handwritten Data

- The best model for On-line handwritten data is HMM model with number of states and symbols set to 15, yielding an accuracy of 77% on test data.

- Confusion matrices for train and test data are as follows:

$$\begin{bmatrix} 80.62 & 6.18 & 13.2 \\ 7.18 & 83.35 & 9.55 \\ 11.67 & 7.23 & 81.1 \end{bmatrix} \begin{bmatrix} 73.23 & 16.77 & 10.1 \\ 10.22 & 78.12 & 11.66 \\ 12.5 & 7.85 & 79.65 \end{bmatrix}$$

Table 1: Classification accuracies of model for on-line handwritten data

nstates, nsymbols	**5**	**10**	**12**	**15**	**20**
$TrainData$	54	73.23	74.17	81.69	77.93
$ValidationData$	31.66	46.67	48.33	71.66	46.66

1.3.2 Speech Data(Isolated)

- The best model for speech data is HMM model with number of states and symbols set to 10, yielding an accuracy of 88.88% on test data.
- Confusion matrices for train and test data are as follows:

$$\begin{bmatrix} 97.5 & 2.2 & 0.3 \\ 0.4 & 97.5 & 2.1 \\ 0.3 & 2.1 & 97.5 \end{bmatrix} \begin{bmatrix} 86.88 & 6.5 & 6.62 \\ 2.24 & 89.12 & 8.64 \\ 1.51 & 7.85 & 90.64 \end{bmatrix}$$

Table 2: Classification accuracies of model for speech isolated data

nstates, nsymbols	**5**	**10**	**15**
$TrainData$	93.33	97.5	95.24
$ValidationData$	87.87	90.90	89.92

2 Dataset2: Two dimensional Artificial Data

2.1 Linearly separable data

2.1.1 Perceptron

The perceptron is an algorithm for supervised learning of binary classifiers. A binary classifier is a function which can decide whether or not an input, represented by a vector of numbers, belongs to some specific class. It is a type of linear classifier, i.e. a classification algorithm that makes its predictions based on a linear predictor function combining a set of weights with the feature vector.

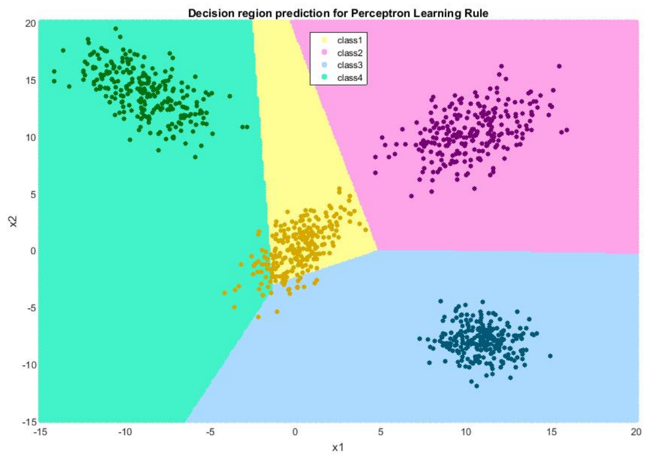

Figure 1: Decision region plot for Perceptron Learning Rule

Table 3: Classification accuracies of the model for different values of η (learning rate)

Learning rate (η)	**1**	**0.1**	**0.01**
$TrainData$	90.7	92.1	95.8
$ValidationData$	89	91.5	94.3

Observation:

- Classification accuracy on test data for the best model is 94.5%
- The decision region boundary of the classifier is linear.
- The best model is chosen for which the accuracy on validation data is maximum which is 94.3% for $\eta = 0.01$
- Confusion matrix(%) for best model on train and test data:

$$\begin{bmatrix} 83.2 & 0 & 5.6 & 11.2 \\ 0 & 100 & 0 & 0 \\ 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 100 \end{bmatrix} \begin{bmatrix} 78 & 0 & 4 & 18 \\ 0 & 100 & 0 & 0 \\ 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 100 \end{bmatrix}$$

2.1.2 Multi-layer feed forward Neural Network

Table 4: Classification accuracies of the model for different values of η (learning rate)

Number of nodes in hidden layers h1 and h2 is 2 and 3 respectively

Learning rate (η)	**0.01**	**0.1**	**5**
$TrainData$	97.1	75.3	74.9
$ValidationData$	97.3	72.7	94.3

Number of nodes in hidden layers h1 and h2 is 3 and 3 respectively

Learning rate (η)	**0.01**	**0.1**	**5**
$TrainData$	100	94. 7	92.4
$ValidationData$	100	92.4	94.7

Observation:

- Classification accuracy on test data for the best model is 100%
- The best model is chosen for which the accuracy on validation data is maximum which is 100% for $\eta = 0.01$

- Confusion matrix(%) for best model on test data:

$$\begin{bmatrix} 100 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 \\ 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 100 \end{bmatrix}$$

2.1.3 Linear kernel based C-SVM

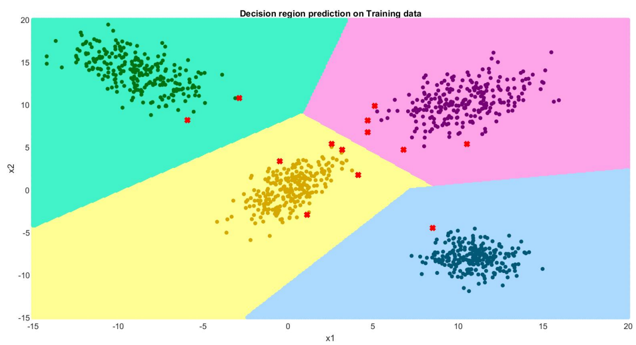

Figure 2: Linear kernel based C-SVM : C = 1, Number of SVs = 13(indicated in red)

Observation:

- Classification accuracy on test data for the best model is 100%.

2.1.4 Polynomial kernel based C-SVM

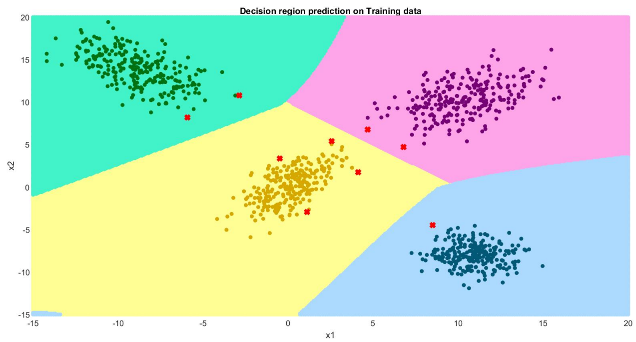

Figure 3: Polynomial kernel based C-SVM : degree = 3, Number of SVs = 9

Observation:

- Classification accuracy on test data for the best model is 100%.
- Parameters for best model : gamma = 1,coef0 = 1,degree = 3,C = 1

2.1.5 Gaussian kernel based C-SVM

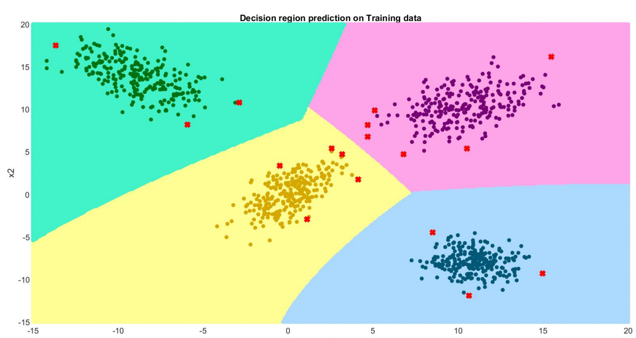

Figure 4: Gaussian kernel based C-SVM : Number of SVs = 17

Observation:

- Classification accuracy on test data for the best model is 100%.
- Parameters for best model : gamma = 0.005, C = 12.

2.1.6 Multi-class logistic regression based classifier using Polynomial basis function

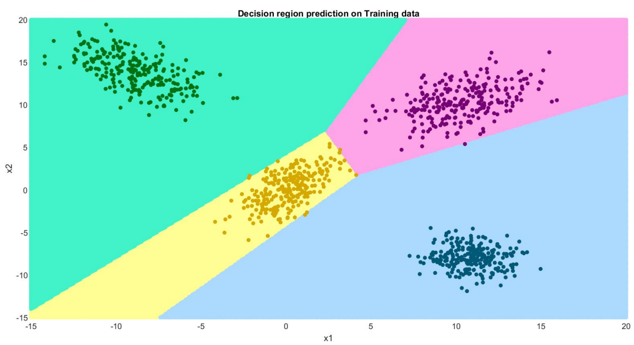

Figure 5: Best Model for Logistic Regression : Degree = 1

Table 5: Classification accuracies of the model for different values of M for $\eta = 1$

M	**1**	**2**	**3**
$TrainData$	100	100	100
$ValidationData$	100	99.6	100

Table 6: Classification accuracies of the model for different values of η for M = 1

η	**1**	**0.1**	**0.01**
$TrainData$	100	100	100
$ValidationData$	100	99.89	99.83

$Observation$:

- Classification accuracy on test data for the best model is 100%.
- The best model is chosen for which the accuracy on validation data is maximum which is 100% for $\eta = 1$.

2.1.7 Multi-class logistic regression based classifier using Gaussian basis function

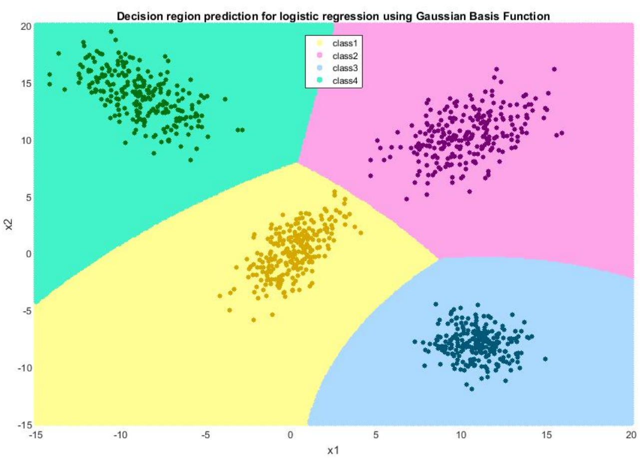

Figure 6: Best Model for Logistic Regression on linearly separable data

Table 7: Classification accuracies of the model for different values of η (learning rate)

Learning rate (η)	**0.1**	**0.01**	**0.001**
$TrainData$	100	100	100
$ValidationData$	99.83	100	100

Observation:

- Classification accuracy on test data for the best model is 100%
- The decision region boundary of the classifier is nonlinear.

- The best model is chosen for which the accuracy on validation data is maximum which is 100% for $\eta = 0.01$

2.2 Nonlinearly separable data

2.2.1 Multi-layer feed forward neural network

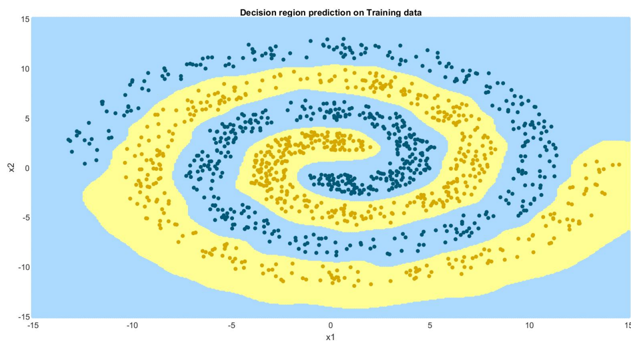

Figure 7: Best Model for multi-layer feed forward neural network on nonlinearly separable data

Table 8: Classification accuracies of the model for different values of η (learning rate)

Number of hidden layer nodes h1 and h2 is 15 and 15 respectively.

Learning rate (η)	**0.1**	**0.01**
$TrainData$	98.1	100
$ValidationData$	96.9	100

Observation:

- Classification accuracy on test data for the best model is 100%
- The decision region boundary of the classifier is nonlinear.
- The best model is chosen for which the accuracy on validation data is maximum which is 100% for $\eta = 0.01$

- Confusion matrix(%) for best model on train data:

2.2.2 Polynomial kernel based C-SVM

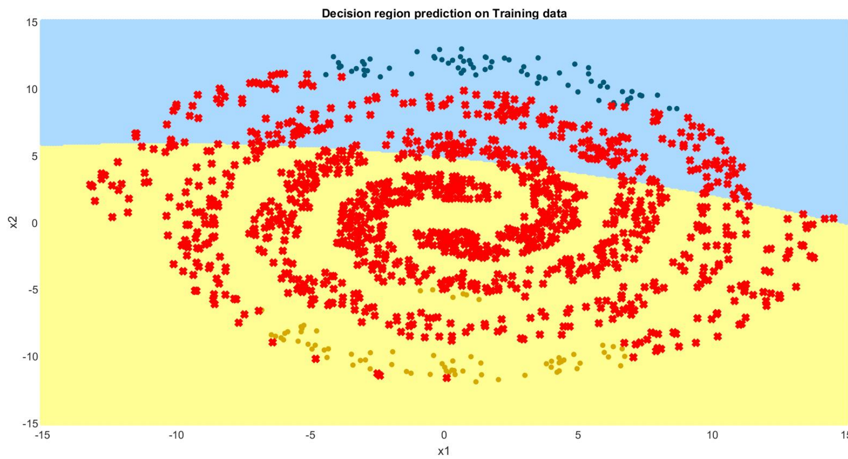

Figure 8: Best model for Polynomial based C-SVM : Degree = 2, Number of SVs = 1161

Table 9: Classification accuracies of the model for different degrees

M	**2**	**3**	**4**	**5**
$TrainData$	56.63	48	49.4	41.5
$ValidationData$	56.9	45.5	51.2	39.8

$Observation$:

- Classification accuracy on test data for the best model is 57.88%.
- The best model is chosen for which the accuracy on validation data is maximum which is 56.9%.
- Parameters for best model : C = 1,degree = 2,gamma = 1,coef0 = 1

- Confusion matrices for train and test data are as follows:

$$\begin{bmatrix} 55.32 & 44.68 \\ 41.52 & 58.48 \end{bmatrix} \begin{bmatrix} 56.9 & 43.1 \\ 43.11 & 56.89 \end{bmatrix}$$

2.2.3 Gaussian kernel based C-SVM

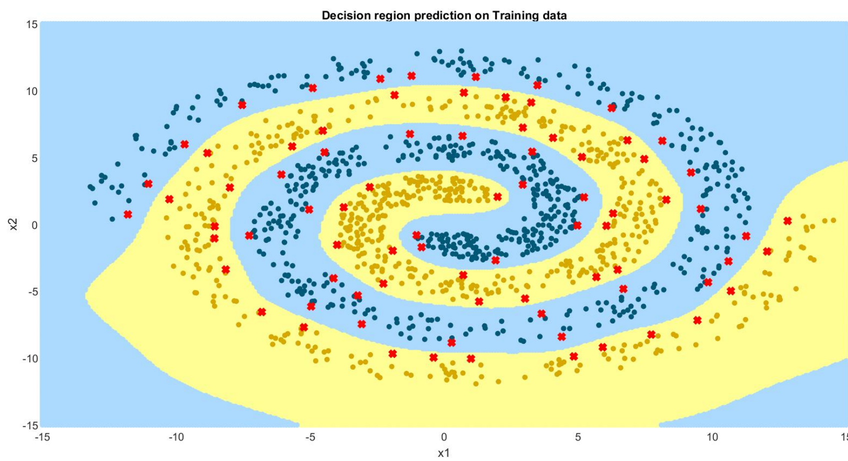

Figure 9: Best model for Gaussian based C-SVM : gamma = 0.1, C = 10, Number of SVs = 80

Table 10: Classification accuracies of the model for different values of gamma

gamma	**1**	**0.5**	**0.1**	**0.05**	**0.01**
$TrainData$	100	100	100	98.8	66.6
$ValidationData$	100	100	100	99.7	65

Observation:

- Classification accuracy on test data for the best model is 100%.
- The parameters corresponding to the best model are : gamma = 0.1 , C = 10 which result in 100% accuracy and the least number of support vectors i.e. 80.

2.2.4 Multi-class logistic regression based classifier using Gaussian basis function

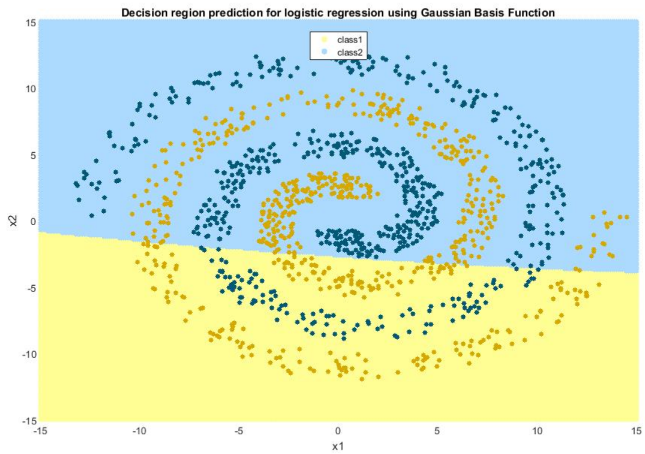

Figure 10: Best Model for Logistic Regression on non-linearly separable data

Table 11: Classification accuracies of the model for different values of η (learning rate)

Learning rate (η)	**0.1**	**0.001**	**0.0001**
$TrainData$	50	50	54.91
$ValidationData$	50	50	58.31

$Observation$:

- Classification accuracy on test data for the best model is 56.92%
- The best model is chosen for which the accuracy on validation data is maximum which is 58.31% for $\eta = 0.0001$
- Confusion matrix(%) for best model on train data:

$$\begin{bmatrix} 32.21 & 67.79 \\ 17.025 & 82.975 \end{bmatrix}$$

- Confusion matrix(%) for best model on test data:

$$\begin{bmatrix} 13.85 & 86.15 \\ 0 & 100 \end{bmatrix}$$

3 Dataset 3: Image Data

3.0.1 Multi-layer feed forward neural network

Table 12: Classification accuracies of the model for different values of nodes in hidden layer h1 and h2

(h1,h2)	**(10,10)**	**(15,12)**	**(12,18)**	**(20,22)**	**(20,25)**
$TrainData$	82.9	50	48.8	89	87.9
$ValidationData$	68.5	65.8	55	70.5	76.5

Observation:

- Classification accuracy on test data for the best model is 77.9%
- The best model is chosen for which the accuracy on validation data is maximum which is 76.5% .
- We observe that as we increase the number of hidden layer nodes, there is an increase in the overall accuracy of the classifier.
- Confusion matrix(%) for best model on train and test data:

$$\begin{bmatrix} 89.32 & 6.18 & 4.5 \\ 1.92 & 95 & 3.08 \\ 3.93 & 16.53 & 79.54 \end{bmatrix} \begin{bmatrix} 80 & 11.43 & 8.57 \\ 8.16 & 71.43 & 20.41 \\ 9.23 & 9.23 & 81.54 \end{bmatrix}$$

3.1 Polynomial kernel based C-SVM

Table 13: Classification accuracies of the model for different degrees

M	**2**	**3**	**4**	**5**	**7**
$TrainData$	100	100	100	100	100
$ValidationData$	57.6	60.7	59.88	58.5	58.66

Observation:

- Classification accuracy on test data for the best model is 69.9%.
- The parameters corresponding to the best model are : degree = 3, gamma = 1 , coef0 = 1, C = 1.
- Number of support vectors = 507.
- Confusion matrix for test data are as follows:

$$\begin{bmatrix} 68.82 & 4.68 & 26.5 \\ 20.52 & 70.23 & 9.25 \\ 20.22 & 9.13 & 70.65 \end{bmatrix}$$

3.2 Gaussian kernel based C-SVM

Table 14: Classification accuracies for different values of gamma

gamma	**1**	**0.01**	**0.005**	**0.001**
$TrainData$	100	100	100	80
$ValidationData$	37.8	59.6	64.9	58.45

Observation:

- Classification accuracy on test data for the best model is 79.9%.

- The parameters corresponding to the best model are : gamma = 0.005 , C = 2.
- Number of support vectors = 625.
- Confusion matrix for test data are as follows:

$$\begin{bmatrix} 78 & 1.5 & 20.5 \\ 16.5 & 80.23 & 3.25 \\ 15.22 & 3.33 & 81.45 \end{bmatrix}$$